PTO/SB/21 (10-07)
Approved for use through 10/31/2007. OMB 0651-0031
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

TRANSMITTAL FORM

(to be used for all correspondence after initial filing)

Total Number of Pages in This Submission	27

Application Number	10/629,448
Filing Date	July 29, 2003
First Named Inventor	Kelkar et. al.
Art Unit	1631
Examiner Name	Loria Clow
Attorney Docket Number	CHA9 2003 0003 US1

ENCLOSURES *(Check all that apply)*

- ☐ Fee Transmittal Form
  - ☐ Fee Attached
- ☐ Amendment/Reply
  - ☐ After Final
  - ☐ Affidavits/declaration(s)
- ☐ Extension of Time Request
- ☐ Express Abandonment Request
- ☐ Information Disclosure Statement
- ☐ Certified Copy of Priority Document(s)
- ☐ Reply to Missing Parts/ Incomplete Application
  - ☐ Reply to Missing Parts under 37 CFR 1.52 or 1.53
- ☐ Drawing(s)
- ☐ Licensing-related Papers
- ☐ Petition
- ☐ Petition to Convert to a Provisional Application
- ☐ Power of Attorney, Revocation Change of Correspondence Address
- ☐ Terminal Disclaimer
- ☐ Request for Refund
- ☐ CD, Number of CD(s) _____
  - ☐ Landscape Table on CD
- ☐ After Allowance Communication to TC
- ☐ Appeal Communication to Board of Appeals and Interferences
- ☑ Appeal Communication to TC **(Appeal Notice, Brief, Reply Brief)**
- ☐ Proprietary Information
- ☐ Status Letter
- ☐ Other Enclosure(s) (please Identify below):

Remarks

SIGNATURE OF APPLICANT, ATTORNEY, OR AGENT

Firm Name	Karl O. Hesse		
Signature	Karl O Hesse		
Printed name	Karl O. Hesse		
Date	October 3, 2007	Reg. No.	25,398

CERTIFICATE OF TRANSMISSION/MAILING

I hereby certify that this correspondence is being facsimile OR EFS-WEB transmitted to the USPTO or deposited with the United States Postal Service with sufficient postage as first class mail in an envelope addressed to: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450 on the date shown below:

Signature	Karl O Hesse		
Typed or printed name	Karl O. Hesse	Date	October 3, 2007

This collection of information is required by 37 CFR 1.5. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.11 and1.14. This collection is estimated to 2 hours to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

If you need assistance in completing the form, call 1-800-PTO-9199 and select option 2.

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Date: October 3, 2007

In re application of:
Kelkar et al
Serial No.: 10/629,448
Filed: July 29, 2003
Group Art Unit: 1631
Examiner: **Loria Clow**
FOR: **Method and Program Product for Discovering Similar Gene Expression Profiles**

I hereby certify that this correspondence is being EFS-Web or facsimile transmitted to the United States Patent and Trademark Office, Fax No. (571) 273-8300

on: 3 Oct 2007 by: KARL O. HESSE.

Karl O Hesse
Signature

3 Oct 2007
Date of Signature

APPEAL BRIEF AND FEE IN SUPPORT OF APPEAL
FROM THE PRIMARY EXAMINER TO THE BOARD OF APPEALS

Assistant Commissioner for Patents
Washington DC 20231

Sir:

Appellants herewith submit an Appeal Brief in support of the appeal to the Board of Patent Appeals and Interferences from the decision dated May 16, 2007 of the Primary Examiner finally rejecting claims 1-6, 10-16 and 20.

FEE

Please charge the fee of $510.00 set by 37 CFR § 41.20(b)(2) for filing a brief in support of an appeal to Deposit Account No. 09-0469. Charge any excess fee or deposit any overpayment to Deposit Account No. 09-0469.

ORAL HEARING

Appellants do not request an Oral Hearing.

(1) Real Party in Interest

The real party in interest in this appeal is International Business Machines Corporation, a New York corporation, assignee of the entire right, title and interest in the claimed invention.

(2) Related Appeals and Interferences

No other appeals or interferences are known to the Appellants, the Appellants' legal representative, or assignee that will directly affect or be directly affected by or have a bearing on the Board's decision in this appeal.

(3) Status of Claims

Claims 1-6, 10-16 and 20 are pending in this application.
Claims 7-9 and 17-19 were canceled after restriction.
The rejection of claims 10-16 and 20 under 35 U.S.C. 101 and the rejection of claims 1-6 under 35 U.S.C. 101 and for new matter is appealed.

(4) Status of Amendments

The amendment filed before final has been entered.
The amendment filed after final has not been entered.

(5) Summary of Invention

The present invention relates to a method and program product operating in a personal computer for clustering genes having potential functional similarity by a comparison of their time varying gene expression profiles.

The method of the invention uses the time and intensity invariant correlation function of the IBM tool to find matches of gene expression profiles using both time and intensity information, which is better at detecting functional similarity than using intensity information alone. The output of Intelligent Miner is a data set of gene expression pairs with the match factor and number of subsets used to compare each pair. A threshold match factor is chosen and genes are listed in clusters by their match fractions. Genes are then removed from all except the cluster with the highest match fraction. Any genes not already in a cluster are added to a cluster which includes a gene that has a highest match fraction with the added gene.

(6) Issues

I. Whether output to a user is a required claim step in order to define an invention, that is a practical application which is useful, concrete and tangible.

II. Whether applicants' teaching of a personal computer with implicit, intrinsic and inherent output means in the specification support claims 1 - 6 without adding new matter.

Grouping of Claims

With respect to the final rejection of claims 1 - 6, 10 - 16 and 20 under 35 U.S.C. 101, the rejected claims are grouped as 2 groups.

Claim 10 is representative of the group I and is related to Issue I.

Claim 1 is representative of the group II and is related to Issue II

(7) Argument

Issue I: Whether output to a user is a required claim step in order to define an invention, that is a practical application which is useful, concrete and tangible.

The Group I Claims

Appellants claim in exemplary claim 10:

10. A program product having computer readable code stored on a recordable media for determining similarity between portions of gene expression profiles comprising:

programmed means for processing a number of gene expression profiles with a similar sequences algorithm that is a time and intensity invariant correlation function to obtain a data set of gene expression pairs and a match fraction for each pair;

programmed means for listing gene expression pairs in clusters by their match fractions;

programmed means for removing a first gene from a cluster when the first gene is also in another cluster which has another gene with a higher match fraction with the first gene than any of the genes in the cluster have with the first gene;

programmed means for repeating the removing step until all genes are listed in only one cluster.

Applicants believe that the specification and claims indeed do describe a method and a program product that produce a result that has substantial and credible utility as required by MPEP 2107 II and that the claims are limited to a narrow practical application in a computer related art.

The Examiner relies on the "New Interim Guidelines" to interpret the requirements of the Federal Courts under the current law to require claiming "output to a user". Applicants believe that the Examiner is mistaken and is applying an interpretation of the definition of the word tangible that is: 1) narrower than appropriate under the current law and is 2) narrower than required under the "New Guidelines".

1) The introduction to the "New Guidelines" states: "These Examination Guidelines ("Guidelines") are based on the USPTO's current understanding of the law and are believed to be fully consistent with binding precedent of the Supreme Court, the Federal Circuit and the Federal Circuit's predecessor courts. These Guidelines do not constitute substantive rulemaking and hence do not have the force and effect of law."

In following the "Guidelines", the Examiner appears to require separate interpretations of the words useful, concrete and tangible.

Applicants' attorney has found no basis in any of the Federal Circuit opinions using these words that imply that these terms are to have separate meanings. They appear to always be used together as synonyms for the concept of being useful and non-abstract. Applicants' attorney has requested that the Examiner provide a citation to a court's requirement that these terms are part of a three pronged test if such is the case in

order to help applicants decide whether to appeal or request continued examination. No citation was provided.

2) Even under the "Guidelines, the Examiners interpretation of the word tangible is unnecessarily narrow.

The "Guidelines" at page 13 recite

"Accordingly, a complete definition of the scope of 35 U.S.C. § 101, reflecting Congressional intent, is that any new and useful process, machine, manufacture or composition of matter under the sun that is made by man is the proper subject matter of a patent. The subject matter courts have found to be outside of, or exceptions to, the four statutory categories of invention is limited to abstract ideas, laws of nature and natural phenomena. While this is easily stated, determining whether an applicant is seeking to patent an abstract idea, a law of nature or a natural phenomenon has proven to be challenging."

Beginning at page 21 the "Guidelines" recite:

"TANGIBLE RESULT"

"The tangible requirement does not necessarily mean that a claim must either be tied to a particular machine or apparatus or must operate to change articles or materials to a different state or thing. However, the tangible requirement does require that the claim must recite more than a § 101 judicial exception, in that the process claim must set forth a practical application of that § 101 judicial exception to produce a real-world result. Benson, 409 U.S. at 71-72, 175 USPQ at 676-77 (invention ineligible because had "no substantial practical application."). "[A]n application of a law of nature or mathematical formula to a ... process may well be deserving of patent protection." Diehr, 450 U.S. at 187, 209 USPQ at 8 (emphasis added); see also Corning, 56 U.S. (15 How.) at 268, 14 L.Ed. 683 ("It is for the discovery or invention of some practical method or means of producing a beneficial result or effect, that a patent is granted . . .").

In other words, the opposite meaning of "tangible" is "abstract." The bare conversion of any binary data as in Gottschalk V. Benson or the bubble sort of any data as in "Warmerdam, 33 F.3d at 1360, 31 USPQ2d at 1759 ("steps of 'locating' a medial axis, and 'creating' a bubble hierarchy . . . describe

nothing more than the manipulation of basic mathematical constructs, the paradigmatic 'abstract idea'")" recited at page 14 of the "Guidelines" are examples of the abstract.

Applicants' process does not convert or process just any data but is limited to useful concrete and non-abstract gene expression profiles in a data base of such profiles. Applicants' process is but one application of many possible applications of the mathematical steps involved in obtaining the useful result.

At page 17 of the "Guidelines we see:

While abstract ideas, natural phenomena, and laws of nature are not eligible for patenting, methods and products employing abstract ideas, natural phenomena, and laws of nature to perform a real-world function may well be. In evaluating whether a claim meets the requirements of section 101, the claim must be considered as a whole to determine whether it is for a particular application of an abstract idea, natural phenomenon, or law of nature, rather than for the abstract idea, natural phenomenon, or law of nature itself.

As is clear from the specification and the claim limitations, applicants' process is limited to a particular practical application and is not an abstract idea, natural phenomenon or a law of nature.

The result is that all of the processed gene expression profiles are each listed in only one cluster. This result of applicants' claims is a very useful, repeatable and non-abstract result which is recognized by those skilled in the medical and computer arts to be of great value and useful, non-abstract and concrete finding of similar gene expression profiles.

PRIOR ART

Applicants note that their claims have not been rejected on prior art yet have been restricted on the ground that there were two groups of claims that required two fields of search. It is not apparent whether relevant prior art patents were considered by the Examiner while examining this application. It is believed that the "Guidelines" on page 10 are helpful in determining both the novelty of applicants' invention and the **usefulness and non-abstract nature** of applicants' the invention.

As evidenced by the references which applicants have attempted to incorporate by reference, but have acquiesced to the Examiners correct requirement to cancel, in addition to applicants teachings in the background art section of their specification, users in the medical profession find great value and usefulness in methods for finding similar gene expression profiles that are tangible and concrete. See for example US Patent 6,406,853 abstract and claims 25, 26 and US Patent 6,436,642 column 26 beginning at line 15.

It is believed that if the rejections under 35 U.S.C. 101 put forth in this application were appropriate, many of the relevant prior art patents in the appropriate fields of search would be found to be invalid. Since they were issued under the guidance of current statutory law and court cases, it must be that the rejections in this application are based upon excessively narrow and untenable interpretation of the current law.

Issue II: Whether applicants' teaching of a personal computer with implicit, intrinsic and inherent output means in the specification support claims 1 - 6 without adding new matter.

The Group II Claims

Appellants claim in exemplary claim 1:

1. A method for determining similarity between portions of gene expression profiles in a computer comprising the steps of:

processing a number of gene expression profiles with a similar sequences algorithm that is a time and intensity invariant correlation function to obtain a data set of gene expression profile pairs and a match fraction for each gene expression profile pair;

listing gene expression profile pairs in clusters by their match fractions;

removing a first gene expression profile from a cluster when another cluster has another gene expression profile with a higher match fraction with the first gene expression profile, unless the another gene expression profile requires a larger number of subsequences to achieve similarity with the first gene expression profile;

repeating the removing step until all gene expression profiles are listed in only one cluster;

providing output of the listing of clusters of gene expression profiles.

Applicants' specification recites: The focal point of the preferred personal computer architecture comprises a processor 51. The processor 51 is connected to a bus 52 which comprises a set of data lines, a set of address lines and a set of control lines. A plurality of I/O devices, memory and storage devices 53-58 and 66 are connected to the bus 52 through separate adapters 59-64 and 67, respectively. For example, the display 54 may be either a CRT or a flat panel display.

It is believed to be well known in the art as exemplified by prior art patents that users in the medical profession receive output from personal computer input/output devices such as applicants teach in their preferred embodiment. Again, applicants refer to US Patent 6,406,853 abstract and claims 25, 26 and US Patent 6,436,642 column 26 beginning at line 15.

It is believed that material that is implicit, intrinsic, or inherent in the application as filed is not new matter.

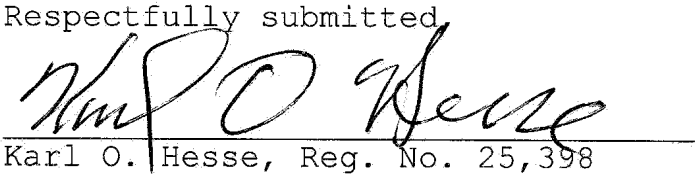
In order to be usable by a user, a personal computer necessarily and constantly exhibits the function of input and output, and such function was recognized as such by those skilled in the art of using personal computers. Therefore applicants' addition of the step of providing such output to satisfy the Examiner's reading of the guidelines was not new matter but is supported in their specification by teachings that are implicit, intrinsic and inherent.

Accordingly it is believed that the claims are clear, statutory and definite and are drawn to a novel and unobvious method and program product for clustering gene expression profiles which result is concrete, tangible and directly useful in drug selection and disease diagnosis.

Request for Relief

Wherefore, Appellants respectfully request that the rejection of pending claims 1 - 6, 10 - 16 and 20 be reversed.

Respectfully submitted,



Date: October 3, 2007

Karl O. Hesse, Reg. No. 25,398
Attorney for Appellants

IBM Corporation
Intellectual Property Law
MG90-201/1
8501 IBM Drive
Charlotte, NC 28262-8563

Land line (704) 895-8241
Cell phone (704) 724-1413
Fax: (704) 594-8307

(8) Appendix

Claims Involved in this Appeal

1. A method for determining similarity between portions of gene expression profiles in a computer comprising the steps of:

processing a number of gene expression profiles with a similar sequences algorithm that is a time and intensity invariant correlation function to obtain a data set of gene expression profile pairs and a match fraction for each gene expression profile pair;

listing gene expression profile pairs in clusters by their match fractions;

removing a first gene expression profile from a cluster when another cluster has another gene expression profile with a higher match fraction with the first gene expression profile, unless the another gene expression profile requires a larger number of subsequences to achieve similarity with the first gene expression profile;

repeating the removing step until all gene expression profiles are listed in only one cluster;

providing output of the listing of clusters of gene expression profiles.

2. A method for determining similarity between portions of gene expression profiles comprising the steps of:

processing a number of gene expression profiles with a similar sequences algorithm that is a time and intensity invariant correlation function to obtain a data set of gene expression pairs and a match fraction for each pair;

listing gene expression pairs in clusters by their match fractions;

removing a first gene from a first cluster when the first gene is also in a second cluster which has another gene with a higher match fraction with the first gene than any of the genes in the first cluster have with the first gene, but;

retaining the first gene in the first cluster and removing the first gene from the second cluster when the difference between the highest match fraction of the first gene with a gene in the first cluster and the highest match fraction of the first gene with a gene in the second cluster is less than a minimum difference threshold and the number of subsequences represented in the similar gene pair having the highest match fraction in the first cluster is higher than the number of subsequences represented in the similar gene pair having the highest match fraction in the second cluster;

repeating the removing step until all genes are listed in only one cluster;

providing output of the listing of clusters of gene expression profiles.

3. A method of determining similarity between portions of gene expression profiles comprising the steps of:

processing data embodying a number of gene expression profiles with a similar sequences algorithm in a computer that is a time and intensity invariant correlation function to obtain a data set of gene expression pairs and a match fraction for each pair;

choosing a threshold match fraction;

listing gene expression pairs in clusters by their match fractions above the threshold;

adding each gene not already in a cluster to a cluster having another gene having a highest match fraction with the each gene without regard of the threshold;

removing a first gene from a cluster when the first gene is also in another cluster which has another gene with a higher match fraction with the first gene than any of the genes in the cluster have with the first gene;

repeating the removing step until all genes are listed in only one cluster;

providing output of the listing of clusters of gene expression profiles.

4. A method for determining similarity between portions of gene expression profiles comprising the steps of:

processing a number of gene expression profiles with a similar sequences algorithm that is a time and intensity invariant correlation function with a computer to obtain a data set of gene expression pairs and a match fraction for each pair;

choosing a threshold match fraction;

listing gene expression pairs in clusters by their match fractions above the threshold;

adding each gene not already in a cluster to a cluster having another gene having a highest match fraction disregarding the threshold with the each gene;

removing a first gene from a first cluster when the first gene is also in a second cluster which has another gene with a higher match fraction with the first gene than any of the genes in the first cluster have with the first gene, but;

retaining the first gene in the first cluster and removing the first gene from the second cluster when the difference between the highest match fraction of the first gene with a gene in the first cluster and the highest match fraction of the first gene with a gene in the second cluster is less than a minimum difference threshold and the number of subsequences represented in the similar gene pair having the highest match fraction in the first cluster is higher than the number of subsequences represented in the similar gene pair having the highest match fraction in the second cluster;

repeating the removing and retaining steps until all genes are listed in only one cluster;

providing output of the listing of clusters of gene expression profiles.

5. A method in a computer for determining similarity between genes comprising the steps of:

listing genes to be compared in a data set by their gene expression profiles;

processing the listed gene expression profiles with a similar sequences algorithm that is a time and intensity invariant correlation function to obtain a data set of gene expression pairs and a match fraction for each pair;

choosing a threshold match fraction;

creating a set G in which to list indices of genes accounted for;

assigning genes i and j to a cluster a if they have a match fraction greater than the threshold;

assigning gene k to the cluster a if it has a match fraction greater than the threshold with either gene i or gene j;

assigning genes k and l to a cluster b if they have a match fraction greater than the threshold and if both gene k and gene l do not have match fractions above the threshold with either gene i or gene j;

repeating the assigning steps until all genes to be compared have been considered;

removing a first gene from a cluster when another cluster has another gene with a higher match fraction with the first gene;

repeating the removing step until all genes are listed in only one cluster;

providing output of the listing of clusters of gene expression profiles.

6. A method in a computer for determining similarity between genes comprising the steps of:

listing genes to be compared in a data set by their gene expression profiles;

processing the listed gene expression profiles with a similar sequences algorithm that is a time and intensity invariant correlation function to obtain a data set of gene expression pairs and a match fraction for each pair;

choosing a threshold match fraction;

creating a set G in which to list indices of genes accounted for;

assigning genes i and j to cluster 1 if they have a match fraction greater than the threshold;

assigning gene k to cluster 1 if it has a match fraction greater than the threshold with either gene i or gene j;

assigning genes k and l to cluster 2 if they have a match fraction greater than the threshold and if both gene k and gene l do not have match fractions above the threshold with either gene i or gene j;

removing a first gene from a cluster when another cluster has another gene with a higher match fraction with the first gene, unless the another gene requires a larger number of subsequences to achieve similarity with the first gene;

repeating the removing step until all genes are listed in only one cluster;

providing output of the listing of clusters of gene expression profiles.

10. A program product having computer readable code stored on a recordable media for determining similarity between portions of gene expression profiles comprising:

programmed means for processing a number of gene expression profiles with a similar sequences algorithm that is a time and intensity invariant correlation function to obtain a data set of gene expression pairs and a match fraction for each pair;

programmed means for listing gene expression pairs in clusters by their match fractions;

programmed means for removing a first gene from a cluster when the first gene is also in another cluster which has another gene with a higher match fraction with the first gene than any of the genes in the cluster have with the first gene;

programmed means for repeating the removing step until all genes are listed in only one cluster.

11. A program product having computer readable code stored on a recordable media for determining similarity between portions of gene expression profiles using output from a similar sequences algorithm that is a time and intensity invariant correlation function comprising:

programmed means for providing a gene expression profile data set as input to programmed means embodying a similar sequences algorithm that is a time and intensity invariant correlation function to obtain a data set of gene expression pairs and a match fraction for each pair as output from the programmed means embodying a similar sequences algorithm;

programmed means for listing the gene expression pairs in clusters by their match fractions;

programmed means for removing a first gene from a cluster when the first gene is also in another cluster which has another gene with a higher match fraction with the first gene than any of the genes in the cluster have with the first gene;

programmed means for repeating the removing step until all genes are listed in only one cluster.

12. A program product having computer readable code stored on a recordable media for determining similarity between portions of gene expression profiles comprising the steps of:

programmed means for processing a number of gene expression profiles with a similar sequences algorithm that is a time and intensity invariant correlation function to obtain a data set of gene expression pairs and a match fraction for each pair;

programmed means for listing gene expression pairs in clusters by their match fractions;

programmed means for removing a first gene from a first cluster when the first gene is also in a second cluster which has another gene with a higher match fraction with the first gene than any of the genes in the first cluster have with the first gene, but;

programmed means for retaining the first gene in the first cluster and removing the first gene from the second cluster when the difference between the highest match fraction of the first gene with a gene in the first cluster and the highest match fraction of the first gene with a gene in the second cluster is less than a minimum difference threshold and the number of subsequences represented in the similar gene pair having the highest match fraction in the first cluster is higher than the number of subsequences represented in the similar gene pair having the highest match fraction in the second cluster;

programmed means for repeating the removing step until all genes are listed in only one cluster.

13. A program product having computer readable code stored on a recordable media for determining similarity between portions of gene expression profiles comprising the steps of:

programmed means for processing a number of gene expression profiles with a similar sequences algorithm that is a time and intensity invariant correlation function to obtain a data set of gene expression pairs and a match fraction for each pair;

programmed means for choosing a threshold match fraction;

programmed means for listing gene expression pairs in clusters by their match fractions above the threshold;

programmed means for adding each gene not already in a cluster to a cluster having another gene having a highest match fraction with the each gene without regard of the threshold;

programmed means for removing a first gene from a cluster when the first gene is also in another cluster which has another gene with a higher match fraction with the first gene than any of the genes in the cluster have with the first gene;

programmed means for repeating the removing step until all genes are listed in only one cluster.

14. A program product having computer readable code stored on a recordable media for determining similarity between portions of gene expression profiles comprising the steps of:

programmed means for processing a number of gene expression profiles with a similar sequences algorithm that is a time and intensity invariant correlation function to obtain a data set of gene expression pairs and a match fraction for each pair;

programmed means for choosing a threshold match fraction;

programmed means for listing gene expression pairs in clusters by their match fractions above the threshold;

programmed means for adding each gene not already in a cluster to a cluster having another gene having a highest match fraction disregarding the threshold with the each gene;

programmed means for removing a first gene from a first cluster when the first gene is also in a second cluster which has another gene with a higher match fraction with the first gene than any of the genes in the first cluster have with the first gene, but;

programmed means for retaining the first gene in the first cluster and removing the first gene from the second cluster when the difference between the highest match fraction of the first gene with a gene in the first cluster and the highest match fraction of the first gene with a gene in the second cluster is less than a minimum difference threshold and the number of subsequences represented in the similar gene pair having the highest match fraction in the first cluster is higher than the number of subsequences represented in the similar gene pair having the highest match fraction in the second cluster;

programmed means for repeating the removing and retaining steps until all genes are listed in only one cluster.

15. A program product having computer readable code stored on a recordable media for determining similarity between genes comprising the steps of:

programmed means for listing genes to be compared by their gene expression profiles;

programmed means for processing the listed gene expression profiles with a similar sequences algorithm that is a time and intensity invariant correlation function to obtain a data set of gene expression pairs and a match fraction for each pair;

programmed means for choosing a threshold match fraction;

programmed means for creating a null set G(0) to hold genes accounted for;

programmed means for assigning genes i and j to cluster 1 if they have a match fraction greater than the threshold;

programmed means for assigning gene k to cluster 1 if it has a match fraction greater than the threshold with either gene i or gene j;

programmed means for assigning genes k and l to cluster 2 if they have a match fraction greater than the threshold and if both gene k and gene l do not have match fractions above the threshold with either gene i or gene j;

programmed means for removing a first gene from a cluster when another cluster has another gene with a higher match fraction with the first gene;

programmed means for repeating the removing step until all genes are listed in only one cluster.

16. A program product having computer readable code stored on a recordable media for determining similarity between genes comprising the steps of:

programmed means for listing genes to be compared by their gene expression profiles;

programmed means for processing the listed gene expression profiles with a similar sequences algorithm that is a time and intensity invariant correlation function to obtain a data set of gene expression pairs and a match fraction for each pair;

programmed means for choosing a threshold match fraction;

programmed means for creating a null set G(0) to hold genes accounted for;

programmed means for assigning genes i and j to cluster 1 if they have a match fraction greater than the threshold;

programmed means for assigning gene k to cluster 1 if it has a match fraction greater than the threshold with either gene i or gene j;

programmed means for assigning genes k and l to cluster 2 if they have a match fraction greater than the threshold and if both gene k and gene l do not have match fractions above the threshold with either gene i or gene j;

programmed means for removing a first gene from a cluster when another cluster has another gene with a higher match fraction with the first gene, unless the another gene requires a larger number of subsequences to achieve similarity with the first gene;

programmed means for repeating the removing step until all genes are listed in only one cluster.

20. In a method of determining similarity between portions of gene expression profiles which includes processing a number of gene expression profiles using a computer with a similar sequences algorithm that is a time and intensity invariant correlation function to obtain a data set of gene expression pairs and a match fraction for each pair, the improvement comprising the steps of:

listing gene expression pairs in clusters by their match fractions;

removing a first gene from a cluster when another cluster has another gene with a higher match fraction with the first gene, unless the another gene requires a larger number of subsequences to achieve similarity with the first gene;

repeating the removing step until all genes are listed in only one cluster;

providing output of the listing of clusters of gene expression profiles.